# 小样本情况下错误先验信息对贝叶斯估计的影响：基于多层模型的研究

郑舒方[1] 张沥今[1] 潘俊豪[1]（通讯作者）
1 中山大学心理学系，广州，510006

**摘要** 在心理学、教育学和组织行为学等领域的研究中，研究者常常会遇到存在嵌套结构的多层数据，如被试可能嵌套于社区、班级、诊所等。如果不考虑数据本身的嵌套结构，可能导致一些统计模型违反其独立性假设，从而对模型参数的估计产生较大的偏差。因此，研究者往往需要利用多层模型解决多层数据中非独立性观测可能带来的问题，但由于客观条件的限制，实际研究中多层数据常常出现水平 1 或水平 2 样本量偏小的情况。基于传统频率学派的极大似然估计方法(Maximum Likelihood, ML)需要依赖大样本，在小样本下容易出现参数估计及模型收敛等方面的问题。贝叶斯估计在小样本的情况下往往具有更大的优势，但与此同时，也更容易受到先验信息主观设置的影响。为了探讨贝叶斯估计中错误先验信息可能带来的负面影响，并与传统方法进行对比，本文基于多层模型，利用蒙特卡洛模拟，研究不同数据类型(因变量分别为连续正态分布数据、连续非正态数据与二分变量的数据)、样本量和组内相关系数(Intraclass Correlation Coefficient, ICC)的情况下，设有不同信息强度和偏差程度的先验信息对贝叶斯估计的影响。总体结果显示，贝叶斯估计中均值严重偏离真值的先验分布会对参数估计带来较大的负面影响，特别是在 ICC 较大，以及群组样本量和先验分布方差较小的情况，且因变量为非正态分布或二分变量时，错误先验信息的负面影响更为明显。本文对多层模型中错误先验信息对贝叶斯估计的影响进行研究并提出建议，希望为贝叶斯估计中先验分布的设置提供一定的理论补充及实证参考。

**关键词** 多层模型；小样本；贝叶斯估计；错误先验信息

# The Influence of Inaccurate Informative Priors on Bayesian Estimation in Small Samples: A Study Based on Multilevel Modeling

Zheng Shufang[1], Zhang Lijin[1], Pan Junhao[1]
1. Department of Psychology, Sun Yat-sen University, China

## Abstract

In the research of psychology, education, and organizational behavior, researchers often encounter multilevel data with hierarchical structures (e.g., participants may cluster within communities, classes, or clinics). Ignoring the hierarchical structures of data may lead to a violation of the independence assumption of some models, resulting in biased parameter estimates. Therefore, researchers often need to conduct multilevel modeling to solve the statistical problems caused by non-independent observations. However, due to the limitation of objective conditions, in real studies, the sample sizes of level 1 and/or level 2 are often small in hierarchical data. Traditional frequentist-based maximum likelihood (ML) approach, which relies on large-sample theory, might lead to problems in parameter estimation and model convergence in multilevel modeling with small samples. In contrast, Bayesian approach is often more advantageous in small samples, but it is also more susceptible to the subjective specification of priors.

To investigate the potentially detrimental effects of inaccurate prior information on Bayesian approaches and compare their performance to the traditional approaches, we conducted a series of simulations under the multilevel model framework with different dependent variable types (i.e., continuous normal, continuous non-normal, and binary dependent variables), sample sizes and intraclass correlation coefficients (ICCs). In sum, the results revealed the devastating impacts of inaccurate prior information on Bayesian estimation, especially in the cases of larger ICC, smaller level 2 sample size, and smaller prior variance. When the dependent variable was non-normal or binary, these negative effects were more obvious. The present study investigated the impacts of inaccurate prior information on Bayesian estimation and provided advice on the specification of priors. We hope that it could contribute to strengthening the theoretical and practical understanding of prior specifications.

**Key words** Multilevel Modeling; Small Samples; Bayesian Modeling; Inaccurate Prior Information

# 1 引言

在心理学、社会学和组织行为学等领域的研究中，常常会涉及到存在嵌套结构的数据，也称为多层数据(Aryee et al., 2012; Holtmann et al., 2016; Hox et al., 2017)。在这类数据中，被试可能来自不同的社区、班级、公司、家庭等，由于同一群组内的个体更倾向于得到一致的结果，在建模中考虑这种内部的一致性是很有必要的(Sutton et al., 2013)。在多层数据中，如果不考虑数据本身的嵌套结构，同一群组内个体存在的依赖性会导致一些统计模型(如，多元线性回归模型)违反独立性假设(Cornfield, 1978; Rutterford et al., 2015; Van Breukelen & Candel, 2012)，从而使模型参数的估计产生较大的偏差(McNeish, 2016)。为了解决由群组内部的非独立性观测所引起的统计方面的问题，研究者往往需要建立多层模型(Multilevel Modeling; Ryu, 2015)。

多层模型的传统估计方法是频率学派的方法(主要为极大似然估计, Maximum Likelihood, ML)，但它往往需要依赖较大的样本量，在群组样本量较小时容易出现模型收敛、参数估计等方面的问题。然而，多层数据的收集往往需要更大的经济成本、更高的收集难度，因此增大群组样本量十分困难(Campbell & Walters, 2014)。特别是在群组随机控制实验(cluster Randomised Controlled Trials, cRCT)中，往往会出现群组样本量过小的问题(McNeish, 2016)。事实上，心理学研究者常常需要通过设计实验研究，通过实验组和控制组的对比，探讨其实验控制或干预是否起到显著的效果。随机控制实验(Randomised Controlled Trials, RCT)正是实验研究设计的黄金标准(Campbell & Walters, 2014; Rutterford et al., 2015)，但许多原因驱使研究者们更倾向于选择对群组而不是个体为单位进行实验。这些原因可能包括：研究受到地理环境的客观限制、希望降低“干预污染”(Treatment Contamination)的风险(Campbell, 2019)，以及希望提高干预的效率、便利性、参与者的依从性，减少伦理方面的争议等(Rutterford et al., 2015)。因此，在心理学、社会学和教育学的研究中，特别是在临床心理学领域，这种 cRCT 的实验设计都是十分常见的(Campbell, 2019; Ribeiro et al., 2018)。

考虑到传统方法在小样本中的局限性，有研究者提出可以利用贝叶斯的方法来解决小样本可能带来的问题(Depaoli & van de Schoot, 2017; Kadane, 2015; McNeish, 2016; van de Schoot et al., 2014)。有模拟研究表明，在多层模型中，相比 ML，贝叶斯方法仅需要更少的群组样本量就可以得到较为准确的参数估计(Hox et al., 2012)。贝叶斯方法的另一优势在于它能结合前人研究的先验信息，提高参数估计的准确性(Depaoli & Clifton, 2015; McNeish, 2016)。然而，由于实证研究和理论依据的缺乏，先验分布的设置很难和真实情况完全一致，而错误

的先验信息可能对模型参数的估计带来较大的影响。

近年来，一些学者开始关注错误先验信息对特定模型可能造成的影响，如多质多法模型(Multitrait-Multimethod, MTMM; Holtmann et al., 2016)、潜发展模型(Depaoli, 2014; Shi & Tong, 2017)和带协变量的 CFA 模型(Multiple Indicators, Multiple Causes Model, MIMIC; Finch & Miller, 2019)等。Holtmann 等人(2016)的研究主要基于 MTMM 的框架，对小样本条件下不同估计方法在多层模型中的表现进行比较。其研究表明，较强的错误先验信息(即先验分布的方差较小)会对模型参数估计带来毁灭性的结果，但对参数设定较弱(即先验分布的方差较大)的错误先验信息时，其估计结果可能优于无信息先验分布的情况。然而，Holtmann 等人(2016)的模型中只考虑了两因子结构的测量模型，没有设置协变量和回归路径，且假设认为水平 2 的样本量较大(50-200)；但在一些多层模型的研究中(如 cRCT)，群组(水平 2)样本量一般小于 50，群组内部(水平 1)的被试数量一般不超过 100(如，van der Putten et al., 2013; Ha et al., 2017; Newton et al., 2018; Shen et al., 2019)。

考虑到 cRCT 研究的普遍性和目前关于错误先验信息的影响研究的局限，本研究将建立以 cRCT 为背景的多层线性模型(McNeish, 2016)，利用蒙特卡洛(Monte Carlo)的方法进行模拟研究。由于实证研究中先验分布的设置存在主观性，本模拟研究将关注多层线性模型中，不同信息强度以及对真实值不同偏差程度的错误先验信息对贝叶斯估计带来的影响，以考察贝叶斯估计对先验分布设置的容错性；同时，对贝叶斯和传统方法(ML)的表现进行对比，以帮助研究者在实际研究中更好地选择合适的估计方法。希望可以为小样本情况下估计方法的选择和贝叶斯方法中先验分布的设置提供借鉴和参考。

# 2 多层线性模型和贝叶斯估计

## 2.1 多层线性模型

多层模型常用来对存在嵌套结构的数据进行分析，以解决由群组内部的非独立性观测所引起的统计方面的问题(Ryu, 2015)。它可以将变量的变异来源分解为不同的水平(如，个体水平、群组水平以及其他更高的水平)，这些变异可以被不同水平的变量共同解释(Heck & Thomas, 2015)，因此能灵活考虑不同水平变量之间的相互作用。

考虑到研究结果的普适性，本研究将主要关注应用最为广泛的多层线性模型，并考虑不同群组间的随机截距和随机斜率效应。这一模型也是标准 cRCT 中的模型(McNeish, 2016)。模型公式和路径图如下(McNeish, 2016)：

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \beta_{3j}X_{3ij} + e_{ij} \tag{1}$$

$$\beta_{0j} = b_{00} + b_{01}W_{1j} + u_{0j} \tag{2}$$

$$\beta_{1j} = b_{10} + b_{11}W_{1j} + u_{1j} \tag{3}$$

$$\beta_{2j} = b_{20} + b_{21}W_{1j} + u_{2j} \tag{4}$$

$$\beta_{3j} = b_{30} + b_{31}W_{1j} + u_{3j} \tag{5}$$

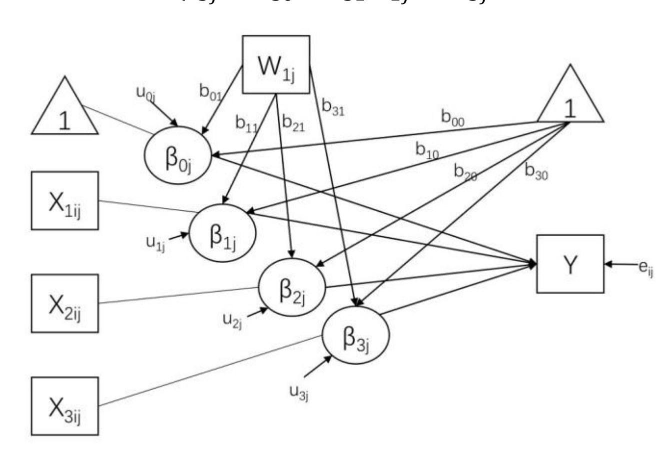



**图 1** 多层线性模型

其中，$W_{1j}$为水平 2 的自变量，在 cRCT 中一般为二分变量，表示实验组或者控制组；$X_{1ij}$-$X_{3ij}$均为水平 1 的自变量，表示被试的性别、年龄、变量前测数据等。$\beta_{0j}$为随机截距，$\beta_{1j}$-$\beta_{3j}$为随机斜率，下标$j$表示这些截距和斜率项在不同群组间可能存在差异。$u_{0j}$-$u_{3j}$分别对应为$\beta_{0j}$-$\beta_{3j}$的残差项，服从均值为 0，方差协方差矩阵为$\Theta$的多元正态分布。

传统上，多层模型常常利用频率学派的方法对未知参数进行估计(Depaoli & Clifton, 2015)。然而，在 ML 的理论框架下，参数的无偏估计需要满足数据渐近符合正态分布的前提，这对样本量要求较高(McNeish, 2016)。在多层模型的分析中，对群组水平(Level 2)的参数进行估计时还需将群组视为整体，因此 ML 方法往往还要求数据有足够的群组数目，从而才能满足渐近理论的假设(Asparouhov & Muthén, 2010)。一般来说，随着多层模型中群组样本量的增多，整体数据更趋向于正态分布(Rutterford et al., 2015)，因此相对于组内样本量，增大群组的数量能更有效地增大实验的检验力(Ribeiro et al., 2018)。较多研究指出，ML 方法在总体样本量较小时的多层模型中表现较为糟糕(McNeish & Stapleton, 2014)，特别是在水平 1 和水平 2 样本量均较小的情况下，传统 ML 方法在多层模型中往往会遇到难以收敛的问题(Depaoli & Clifton, 2015; Hox & Maas, 2001; Hox et al., 2010; McNeish, 2016; Schoeneberger, 2015)，而且可能会出现参数估计结果不可靠、不稳定的情况(Asparouhov & Muthén, 2010; Hox & Maas, 2001; McNeish, 2016)。

## 2.2 贝叶斯估计

为了解决在群组样本量较小的多层模型中传统的频率学派方法在模型收敛、参数估计等方面的问题，研究者可以利用贝叶斯的方法来对模型进行估计(Depaoli & van de Schoot, 2017; Kadane, 2015; McNeish, 2016; van de Schoot et al., 2014)。一般来说，在小样本情况下，贝叶斯估计在模型收敛、参数估计等方面的表现比传统方法更优(Lee & Song, 2004; Muthén & Asparouhov, 2012)，特别是多层模型中群组样本量较小的情况下(Asparouhov & Muthén, 2010; Hox et al., 2012; Muthén & Asparouhov, 2012)。简单来说，贝叶斯方法认为未知参数是随机变量，可以利用先验信息和似然函数得到的后验分布来反映参数在不同取值上的可能性(Lynch, 2007)。它和传统频率学派的基础理论框架不同，不依赖于频率学派的大样本渐近理论(McNeish, 2016)。在贝叶斯估计中，参数取值概率的后验分布可以表示为(Lynch, 2007)：

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{6}$$

其中，$y$表示观测变量向量，θ 对应表示未知参数向量。$p(\theta|y)$表示取值概率的后验分布，它同时取决于先验信息$p(\theta)$和数据似然函数$p(y|\theta)$(Lynch, 2007)。一般来说，当样本量较大时，后验分布主要取决于似然函数，而随着样本量的减小，先验信息对参数后验估计的影响也逐渐增大。因此，在小样本的情况下，对于先验分布的设置需要更加慎重(Holtmann et al., 2016; McNeish, 2016)。

贝叶斯方法不需要对参数的抽样分布做出假设，也不需要通过复杂的计算来得到参数估计的标准误，所以能很好地克服传统方法在小样本情况模型难以达到收敛的问题(Depaoli & Clifton, 2015; Levy & Choi, 2013)。贝叶斯方法的另一优势在于它能结合前人研究的先验信息，提高参数估计的准确性(Depaoli & Clifton, 2015; McNeish, 2016)。然而，先验分布的设置存在一定的主观性，且由于实证研究和理论依据的缺乏，很可能出现先验分布的设置和真实情况不一致的情况，即可能出现错误先验信息的情况。

为了探究贝叶斯估计下，不同信息强度的先验分布(即先验分布方差大小的不同)可能带来的影响，一些研究者对不同情境下的模型进行了模拟研究(Depaoli & Clifton, 2015)。如，McNeish(2016)曾以 cRCT 为研究背景，建立多层模型，利用不同的估计方法对模型参数进行估计和比较，包括传统频率学派的 ML、限制性极大似然估计(Restricted Maximum Likelihood, REML)和贝叶斯估计；并在贝叶斯方法下，比较先验分布的不同信息强度对参数估计的影响。其模拟研究结果表明，先验分布中不同信息强度的设置对模型参数估计的影响较大，特别是在总体样本量较小、因变量为分类数据的情况下(Lynch, 2007; McNeish, 2016)。然而，和许多模拟研究一样，McNeish(2016)对于先验分布的影响研究，是基于其先验分布能提供正确先验的假设，而没有考虑错误先验信息可能带来的影响。一般来说，较强的错误

先验信息会对模型参数估计带来毁灭性的结果，特别是在样本量较小的情况下。但也有研究指出，对参数设定较弱(即先验分布的方差较大)的错误先验信息时，其估计结果可能优于无信息先验分布的情况(Holtmann et al., 2016)。目前，对于贝叶斯估计中错误先验信息的影响仍有许多不明确的地方，需要更多相关研究来加深研究者们的认识(Finch & Miller, 2019; Holtmann et al., 2016)。因此，本文将基于应用较为广泛的多层线性模型，展开一系列模拟研究，以探究不同数据条件下错误先验信息对贝叶斯估计的影响。

# 3 模拟研究

## 3.1 研究设计

基于图 1 的多层线性模型[1]，本研究考虑以下因素的影响：不同样本量(包括水平 1 和水平 2 样本量)、组内相关系数(Intraclass Correlation Coefficient, ICC)以及估计方法。本文利用 R 软件进行模拟数据的生成，并利用 M*plus* 8.0(Muthén & Muthén, 1998-2017)进行模型分析。为简化模型，本研究中因变量服从连续正态分布。

**样本量**

以“cluster Randomised Controlled Trials”和“multilevel”为关键词，在 Web of Science 上对近 5 年(2015-2019 年)的核心期刊论文进行检索及初步筛选，得到 95 篇相关文献。利用摘要、全文获取等方法，对有报告样本量的文献中的群组数量和群组内被试的平均数量进行编码，得到其样本量设置情况。如图 2 所示，近 5 年 cRCT 研究中的群组数量主要集中于 10-40 之间，而群组内被试的平均数量大多小于 100。因此，结合文献检索、编码的结果，以及 Depaoli 和 Clifton(2015)的建议，本研究中水平 2 样本量分别设为 20，30，40，50；水平 1 的样本量分别为：30，60，150。

[1] 其中，$X_{1ij}$作为连续变量，而$X_{2ij}$、$X_{3ij}$、$W_{1j}$均视作二分变量；各斜率$\beta$的截距项($b_{00}$-$b_{30}$)的真值分别设为 0，2.3，-0.25，-0.75；$W_{1j}$对各斜率$\beta$的作用($b_{01}$-$b_{31}$)真值分别设为 1.5，0.12，-0.05，-0.75。

**图 2** 2015-2019 年 cRCT 研究中的样本量统计

**ICC**

在多层模型的研究中，研究者往往还会关注不同水平的 ICC 可能对参数估计造成的影响。ICC 表示变量的总变异可以被水平 2 的变异解释的比例，其计算公式为$\mathrm{ICC} = \sigma_B^2/(\sigma_B^2 + \sigma_W^2)$，其中$\sigma_W^2$和$\sigma_B^2$分别表示水平 1 和水平 2 的方差。ICC 越大，变量在组间的差异相对于组内差异越大。多项前人研究表明，当 ICC 较小时，传统频率学派的方法可能出现参数估计以及模型收敛方面的问题(Hox & Maas, 2001; Preacher et al., 2011; Koch et al., 2015)，而贝叶斯估计的方法可能可以在 ICC 较小时仍能得到准确的参数估计(Hox et al., 2012; Muthén & Asparouhov, 2012)。有模拟研究的结果表明，ICC 越大，无信息先验的贝叶斯方法对水平 2 参数的估计偏差可能反而越大(Depaoli & Clifton, 2015)。另外，在多层 MIMIC 中，对于水平 2 参数的估计偏差会随着 ICC 的增大而增大，检验力随着 ICC 的增大而减小(Cao et al., 2019; Finch & French, 2011)。Fang 等人(2019)指出，未来研究可以进一步探讨在更多多层模型中，ICC 对于贝叶斯估计及其他估计方法表现的影响。

一般来说，当 ICC ≥ 0.059 时，需要考虑被试间的差异，使用多层模型对数据进行分析(Cohen, 1988)。前人关于多层模型的模拟研究中，主要关注的 ICC 介于 0.05 到 0.2 之间(e.g., Depaoli & Clifton, 2015; Hox et al., 2010; Lüdtke et al., 2011; Preacher et al., 2011)，因此本研究中的 ICC 也将分为 0.05、0.1 和 0.2 三个水平。为了使得 ICC 的水平分别为 0.05、0.1 和 0.2，模型中各个水平变量的方差和残差方差随之发生变化(Depaoli & Clifton, 2015)。

**估计方法**

模拟研究使用的估计方法为传统频率学派的 ML[2]估计以及贝叶斯方法。贝叶斯方法分为无信息先验和有信息先验的情况。在有信息先验的贝叶斯估计中，针对回归系数设置先验分布 $N(\mu, \sigma^2)$。μ 分 5 个水平，即无偏；向左/右偏移回归系数真值 1 个标准差(Standard Deviation, *SD*)；以及向左/右偏移真值 3*SD*。$\sigma^2$ 分 3 个水平，分别为对应回归系数真值的 10%、20%、50% (Depaoli, 2014)。以$b_{01}$(真值为 1.5)为例，其 15 种不同偏差程度及信息强度的先验分布设置如图 3 所示。加上无信息先验的情况(按照 M*plus* 默认的设置，回归系数的先验分布均值为 0，方差为无穷大)，贝叶斯方法有 1+5*3=16 种先验分布的设置。再加上传统的估计方法(ML)，需要进行 17 种估计方法的比较。

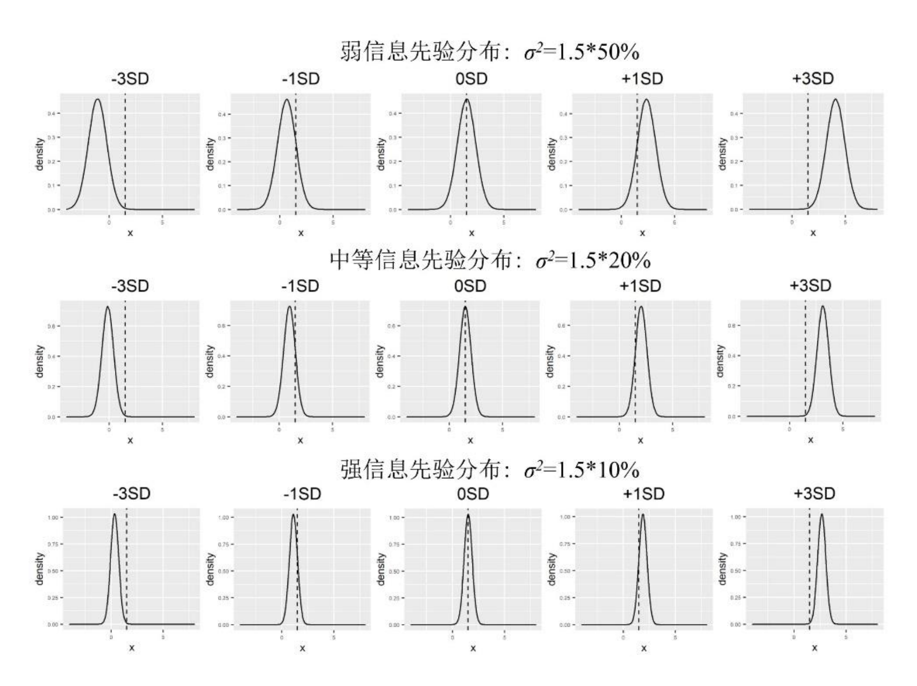



**图 3** 不同偏差程度及信息强度的先验分布设置(以 $b_{01}$ 为例)

综上，本研究包括 3*4*3*17=612 种模拟条件的组合。为了减少模型估计过程中的偶然性因素带来的偏差，本研究在每个条件下均生成 100 个重复样本进行估计。

### 3.2 结果评估指标

为了对不同估计方法的参数估计表现进行评估，本文将首先对各条件下的收敛率进行检查，再根据达到收敛的结果，对 5 个方面指标进行整理、对比和分析。结果评估的指标包括系数估计的相对偏差(Relative Bias)、标准误比例、均方误差(Mean Square Error)、95%置信区间(Confidence Interval)或可信区间(Credible Interval)覆盖率(95% CI 覆盖率)和检验力(Power)。

### 3.3 结果

[2] M*plus* 默认对多层模型采用稳健极大似然估计(Robust Maximum Likelihood Estimator, MLR)

多层线性模型较为简单，因此在本研究考虑的情况中，模型均能得到收敛。本文只针对回归系数进行不同先验分布的设置，因此同样只对回归系数的结果进行评估和分析。为了探究不同结果指标受到不同因素的影响程度及其交互作用，本文还对每个重复样本中各回归系数的结果评估指标进行方差分析(其中，考虑到各参数相对偏差可能出现正负抵消的情况，在方差分析中对相对偏差取绝对值后再进行分析)。如表 1 所示，总体来说，各结果评估指标在不同程度上受到 ICC、水平 1 及水平 2 样本量、先验分布的方差大小、均值偏移真值的程度的影响，且不同因素之间可能相互影响。

在 cRCT 中，研究者往往更关注水平 2 的自变量 $W_{1j}$ 对因变量的影响$b_{01}$。因此，为了使得结果的呈现和分析更清晰，以下结果均将以$b_{01}$为例，分别对其参数估计偏差、标准误比例等评估指标进行分析。



**相对偏差**

相对偏差指模型达到收敛的次数中，模型参数估计得到的均值相对于参数的真值的偏差，其计算公式为$Relative\ Bias = (\hat{\theta} - \theta)/\theta$。其中，$\hat{\theta}$表示模型达到收敛的次数中，参数的平均估计值；$\theta$表示对应的参数真值。相对偏差在±10%以内是可以接受的，可以视为无偏的估计(Flora & Curran, 2004)。

总体来说，各参数估计的相对偏差变化呈现的规律大体一致。以$b_{01}$估计的相对偏差结果为例(图 4)，ML 和无信息先验的贝叶斯估计表现较为接近，均不容易受到不同 ICC 水平的影响。对于有信息先验的贝叶斯方法，其参数估计的偏差会受到水平 2 样本量、先验分布的方差大小以及其均值偏移真值的程度的共同影响，而水平 1 样本量的对相对偏差的影响并不明显。一般来说，随着水平 2 样本量和先验分布方差的增大，先验分布均值对真值的偏移程度对相对偏差的影响也逐渐减小。对于 cRCT 研究中最受关注的实验干预对因变量产生的效应($b_{01}$)，其先验设置中的错误信息对相对偏差的负面影响会随着 ICC 的减小而逐渐减小。但 ICC 的这种影响趋势在其他回归系数中并不明显。

**标准误比例**

标准误比例是模型达到收敛的次数中，标准误估计的平均值(Standard Error, SE)和 SD 的比例。标准误估计越准确，二者的比例应越趋近于 1。一般来说，标准误比例应介于 0.9 和 1.1 之间(Cham et al., 2012)。

$b_{01}$标准误比例的结果如图 5 所示。总体来看，有信息先验的贝叶斯估计往往高估标准误比例的大小，且远远超过可以被接受的范围。随着先验分布方差的增大，先验信息的强度逐渐减弱，标准误比例也逐渐减小并向 1 靠拢。水平 2 样本量的增大也能较大程度地减小有

信息先验的贝叶斯估计中标准误比例的大小。当先验分布均值偏移参数本身真值的程度较大时，其标准误比例往往高于同等条件下正确先验信息的情况。随着 ICC 的减小，错误先验信息对标准误比例的影响也逐渐减小。

**均方误差**

均方误差是模型达到收敛的次数中，参数估计的误差的均值。其计算公式为 $Mean\ Square\ Error = \sum(\hat{\theta}_i - \theta_i)^2$ / 收敛次数。其中，$\hat{\theta}_i$表示模型达到收敛时，第 i 个重复样本得到的参数估计值；$\theta_i$表示对应的参数真值。均方误差越小，参数估计越准确。

如图 6 所示，在小样本的情况下，ML 和无信息先验的贝叶斯估计更倾向于得到的较大的均方误差。正确先验信息的贝叶斯估计在均方误差方面表现最优，而错误先验信息会大大增加贝叶斯方法对参数估计的均方误差。随着 ICC 的减小以及水平 1、水平 2 样本量的增大，先验分布均值偏移真值的程度对均方误差影响也逐渐减小。相比水平 2 样本量的大小，水平 1 样本量对于均方误差的影响较弱。随着水平 2 样本量的增大，ICC 的增大对于错误先验信息负面影响的增强作用也逐渐减小。另外，先验分布的方差越大，错误先验信息对均方误差带来的负面影响也越弱。

**覆盖率**

在模型收敛的重复样本中，参数真值落在 95%置信区间或可信区间的概率即为 95% CI 覆盖率。简单来说，频率学派下通过计算得到的置信区间需要依赖渐进理论的前提假设，而贝叶斯学派下的可信区间直接根据参数后验分布的百分位值得到，因此不需要假设参数渐进服从某一分布。若区间估计准确，95%区间覆盖率应接近 0.95。一般来说，95% CI 覆盖率大于 0.9 是可以被接受的(Collins et al., 2001)。

如图 7 所示，总体来说，随着水平 2 样本量的增大，参数估计的 95% CI 覆盖率逐渐增大，但水平 1 样本量的增大对于 95% CI 覆盖率的影响并不明显。对于有信息先验的贝叶斯估计，一般来说，先验分布均值偏离真值的程度越大，95% CI 覆盖率越低。当先验分布的均值偏离参数真值的程度不超过 1 个标准差时，参数估计的 95% CI 覆盖率均处于可以被接受的范围(大于 90%)；而先验分布的均值偏离参数真值的程度较大时(偏离真值 3 个标准差)，参数估计的 95% CI 覆盖率会受到较大程度的影响。随着 ICC 的减小以及水平 2 样本量的增大，错误先验信息对 95% CI 覆盖率的负面影响逐渐减小；然而，ICC 的这种影响趋势在其他回归系数中并不明显。

**检验力**

检验力指零假设为假的情况下，假设检验成功地拒绝零假设的概率。即当参数的真值不

为 0 的情况下，在模型达到收敛的次数中，该参数估计得到的 95%置信区间(Confidence Interval)或可信区间(Credible Interval)不包括 0 的比例。检验力大于 0.8 一般是可以被接受的(Cohen, 1988; Muthén & Muthén, 2002)。

如图 8 所示，当样本量较小、ICC 较大时，ML 和无信息先验的贝叶斯估计对参数的检验力较低。而对于提供正确先验信息的贝叶斯估计，即使在样本量较小的情况下，其参数估计的检验力仍处于可以被接受的范围(检验力高于 0.8)。低估参数效应的错误先验信息往往降低参数估计的检验力，高估参数效应的错误先验信息对应的检验力也较高，但同时可能导致较高的一类错误率。随着水平 2 样本量的增大以及 ICC 的减小，错误先验信息对检验力的影响也逐渐减小。

**小结**

总的来看，对贝叶斯估计设置正确的有信息先验时，其各项评估指标的表现最优，特别是参数估计的偏差方面。无信息先验的贝叶斯估计和 ML 方法在参数估计上的表现呈现出较为相似的规律，其在均方误差和检验力方面表现不佳，特别是在样本量较小、ICC 较大的情况下。而先验分布均值对参数真值存在偏离时，较大的 ICC 和较小的水平 2 样本量均会增大错误先验信息对参数估计带来的负面影响。先验分布的信息强度越强，其错误先验信息对参数估计的影响也越大。

**表 1** 方差分析 partial $\eta^2$ 结果

| 效应 | 估计偏差 | | | 标准误比例 | | | 均方误差 | | | 95% CI 覆盖率 | | | 检验力 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ML** | **BD** | **INF** | **ML** | **BD** | **INF** | **ML** | **BD** | **INF** | **ML** | **BD** | **INF** | **ML** | **BD** | **INF** |
| ICC | 0.004 | 0.005 | 0.001 | 0.016 | 0.022 | 0.024 | 0.014 | 0.013 | **0.135** | 0.002 | 0.001 | 0.008 | 0.000 | 0.007 | 0.001 |
| L2 | **0.092** | **0.093** | **0.127** | 0.028 | **0.152** | **0.925** | **0.160** | **0.159** | **0.153** | 0.004 | 0.002 | 0.036 | **0.124** | **0.284** | **0.067** |
| L1 | 0.036 | 0.037 | 0.020 | 0.007 | 0.010 | **0.690** | **0.064** | **0.064** | 0.014 | 0.001 | 0.002 | 0.004 | **0.065** | **0.085** | 0.019 |
| mag | - | - | **0.891** | - | - | **0.496** | - | - | **0.540** | - | - | **0.765** | - | - | **0.205** |
| strength | - | - | **0.386** | - | - | **0.956** | - | - | 0.058 | - | - | **0.077** | - | - | **0.108** |
| ICC * L2 | 0.001 | 0.001 | 0.000 | 0.015 | 0.019 | **0.103** | 0.002 | 0.002 | 0.015 | 0.002 | 0.001 | 0.000 | 0.006 | 0.010 | 0.002 |
| ICC * L1 | 0.001 | 0.001 | 0.000 | 0.009 | 0.011 | **0.136** | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.000 | 0.001 | 0.002 | 0.000 |
| ICC * mag | - | - | 0.000 | - | - | 0.002 | - | - | **0.169** | - | - | 0.016 | - | - | 0.006 |
| ICC * strength | - | - | 0.000 | - | - | 0.000 | - | - | 0.008 | - | - | 0.001 | - | - | 0.000 |
| L2 * L1 | 0.001 | 0.001 | 0.000 | 0.027 | 0.042 | **0.462** | 0.007 | 0.007 | 0.001 | 0.001 | 0.002 | 0.000 | 0.003 | 0.004 | 0.001 |
| L2 * mag | - | - | **0.167** | - | - | **0.256** | - | - | **0.185** | - | - | **0.095** | - | - | 0.001 |
| L2 * strength | - | - | 0.013 | - | - | **0.471** | - | - | 0.010 | - | - | 0.002 | - | - | 0.006 |
| L1 * mag | - | - | 0.024 | - | - | 0.007 | - | - | 0.010 | - | - | 0.008 | - | - | 0.000 |
| L1 * strength | - | - | 0.001 | - | - | **0.187** | - | - | 0.001 | - | - | 0.000 | - | - | 0.002 |
| mag * strength | - | - | **0.070** | - | - | **0.083** | - | - | 0.001 | - | - | **0.209** | - | - | 0.013 |
| ICC * L2 * L1 | 0.002 | 0.002 | 0.000 | 0.031 | 0.030 | **0.098** | 0.002 | 0.002 | 0.000 | 0.002 | 0.004 | 0.001 | 0.004 | 0.004 | 0.001 |
| ICC * L2 * mag | - | - | 0.000 | - | - | 0.007 | - | - | 0.027 | - | - | 0.001 | - | - | 0.001 |
| ICC * L2 * strength | - | - | 0.000 | - | - | 0.011 | - | - | 0.007 | - | - | 0.000 | - | - | 0.001 |
| ICC * L1 * mag | - | - | 0.001 | - | - | 0.005 | - | - | 0.001 | - | - | 0.000 | - | - | 0.000 |
| ICC * L1 * strength | - | - | 0.000 | - | - | 0.011 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |
| ICC * mag * strength | - | - | 0.000 | - | - | 0.000 | - | - | 0.006 | - | - | 0.002 | - | - | 0.000 |
| L2 * L1 * mag | - | - | 0.001 | - | - | 0.007 | - | - | 0.004 | - | - | 0.001 | - | - | 0.001 |

| 效应 | 估计偏差 | | | 标准误比例 | | | 均方误差 | | | 95% CI 覆盖率 | | | 检验力 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ML | BD | INF | ML | BD | INF | ML | BD | INF | ML | BD | INF | ML | BD | INF |
| L2 * L1 * strength | - | - | 0.000 | - | - | 0.024 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |
| L2 * mag * strength | - | - | 0.014 | - | - | 0.027 | - | - | 0.005 | - | - | 0.002 | - | - | 0.007 |
| L1 * mag * strength | - | - | 0.001 | - | - | 0.001 | - | - | 0.000 | - | - | 0.001 | - | - | 0.001 |
| ICC * L2 * L1 * mag | - | - | 0.000 | - | - | 0.010 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |
| ICC * L2 * L1 * strength | - | - | 0.000 | - | - | 0.009 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |
| ICC * L2 * mag * strength | - | - | 0.000 | - | - | 0.001 | - | - | 0.011 | - | - | 0.001 | - | - | 0.000 |
| ICC * L1 * mag * strength | - | - | 0.000 | - | - | 0.001 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |
| L2 * L1 * mag * strength | - | - | 0.000 | - | - | 0.001 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |
| ICC * L2 * L1 * mag * strength | - | - | 0.000 | - | - | 0.001 | - | - | 0.000 | - | - | 0.000 | - | - | 0.000 |

**注：**①ML 为极大似然估计；BD 为无信息先验的贝叶斯估计；INF 为有信息先验的贝叶斯估计。②ICC 为组内相关系数；L1 和 L2 分别表示水平 1 和水平 2 样本量；mag 表示先验分布均值偏移真值的程度；strength 表示先验分布的方差大小。③加粗表示中等程度以上的效应量，即 partial $\eta^2$ 大于 0.06。

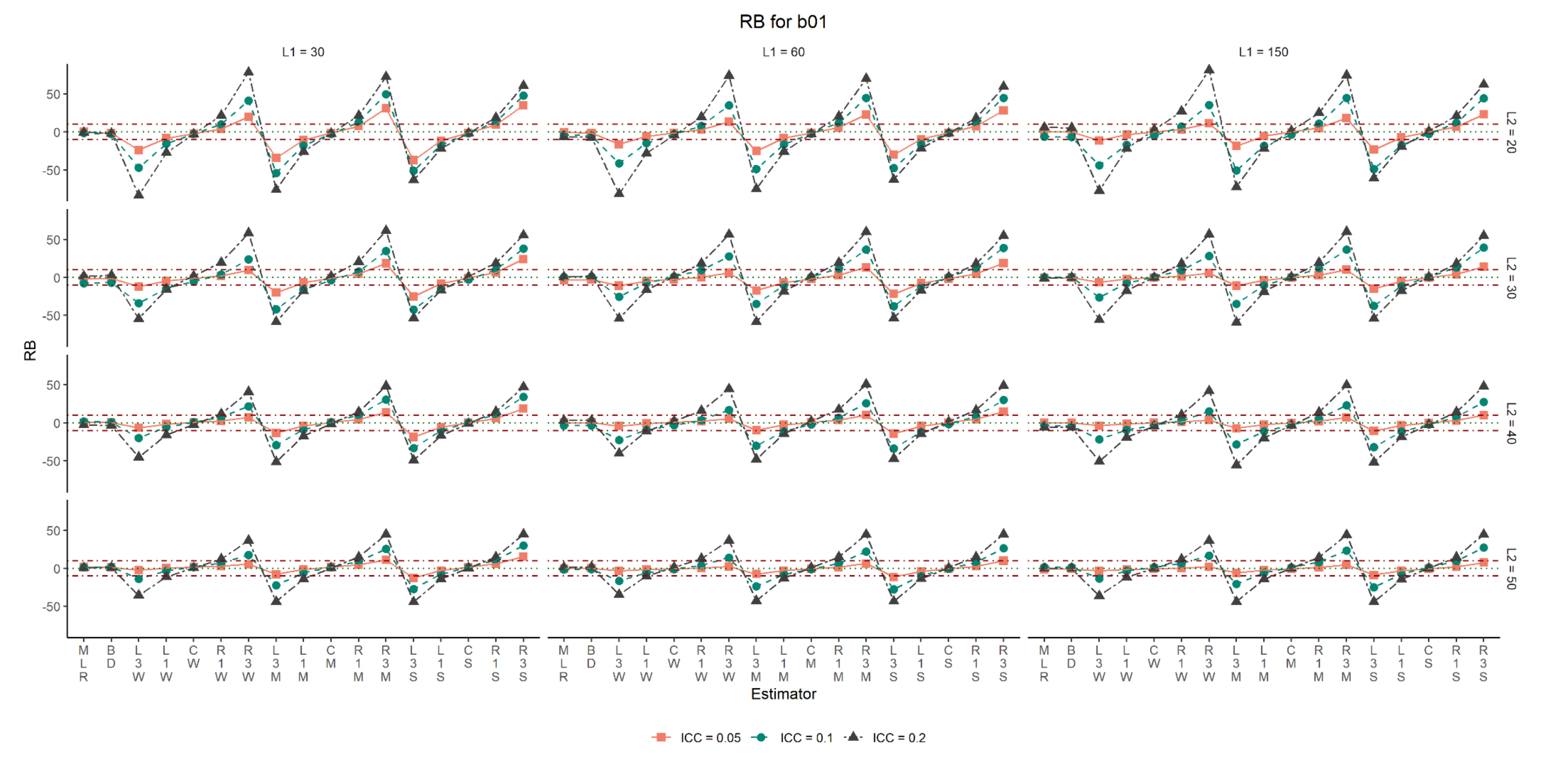



**图 4** 相对偏差结果示例

**注：**①RB 为相对偏差；BD 为无信息先验的贝叶斯估计；L3 为先验分布均值向左偏离真值 3 个标准差；L1 向左偏离真值 1 个标准差；C 表示提供正确先验信息；R1 向右偏离真值 1 个标准差；R3 向右偏离真值 3 个标准差。W 表示先验信息分布的方差较大(真值的 50%)，M 表示先验信息分布的方差中等(真值的 20%)，S 表示先验信息分布的方差较小(真值的 10%)。②图中绿色虚线表示相对偏差恰好为 0 的参照；上下两条红色点划线对应的纵坐标分别为-10 和 10，表示可接受的相对偏差的范围。

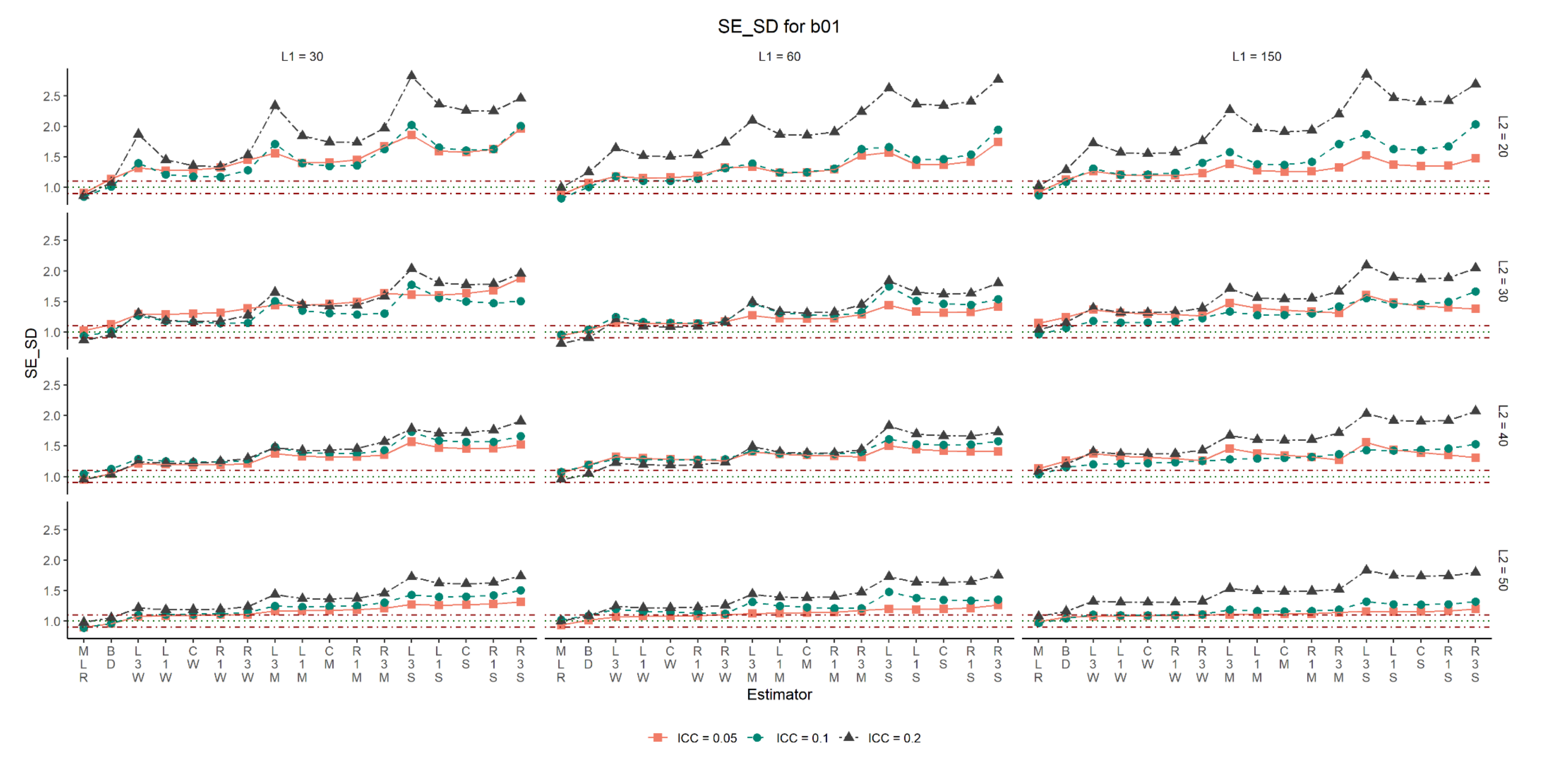



**图 5** 标准误比例结果示例

**注：**①SE_SD 为标准误比例；BD 为无信息先验的贝叶斯估计；L3 为先验分布均值向左偏离真值 3 个标准差；L1 向左偏离真值 1 个标准差；C 表示提供正确先验信息；R1 向右偏离真值 1 个标准差；R3 向右偏离真值 3 个标准差。W 表示先验信息分布的方差较大(真值的 50%)，M 表示先验信息分布的方差中等(真值的 20%)，S 表示先验信息分布的方差较小(真值的 10%)。②图中绿色虚线表示标准误比例恰好为 1 的参照；上下两条红色点划线对应的纵坐标分别为 0.9 和 1.1，表示可接受的标准误比例的范围。

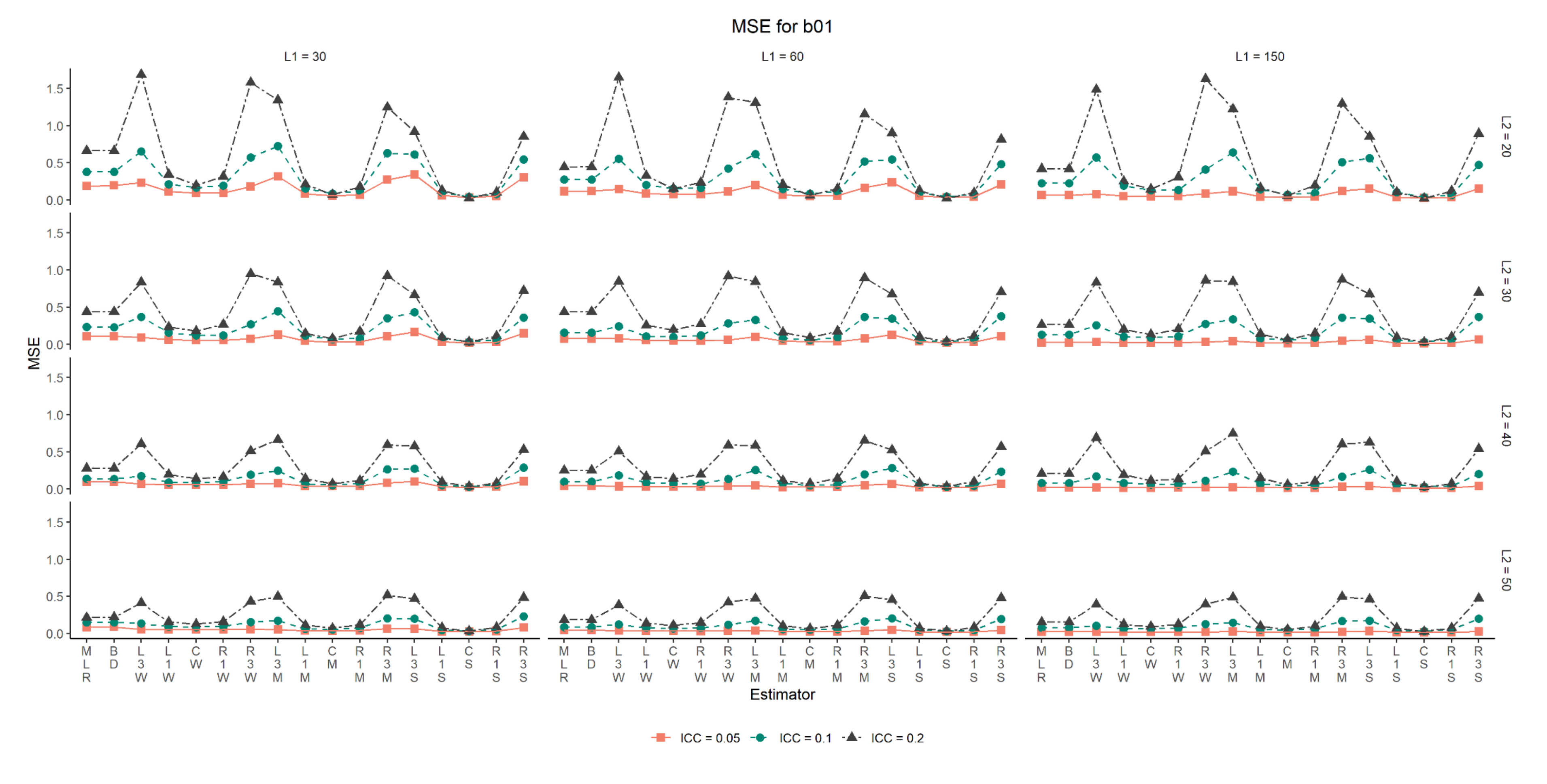



**图 6** 均方误差结果示例

**注：**MSE 为均方误差；BD 为无信息先验的贝叶斯估计；L3 为先验分布均值向左偏离真值 3 个标准差；L1 向左偏离真值 1 个标准差；C 表示提供正确先验信息；R1 向右偏离真值 1 个标准差；R3 向右偏离真值 3 个标准差。W 表示先验信息分布的方差较大(真值的 50%)，M 表示先验信息分布的方差中等(真值的 20%)，S 表示先验信息分布的方差较小(真值的 10%)。

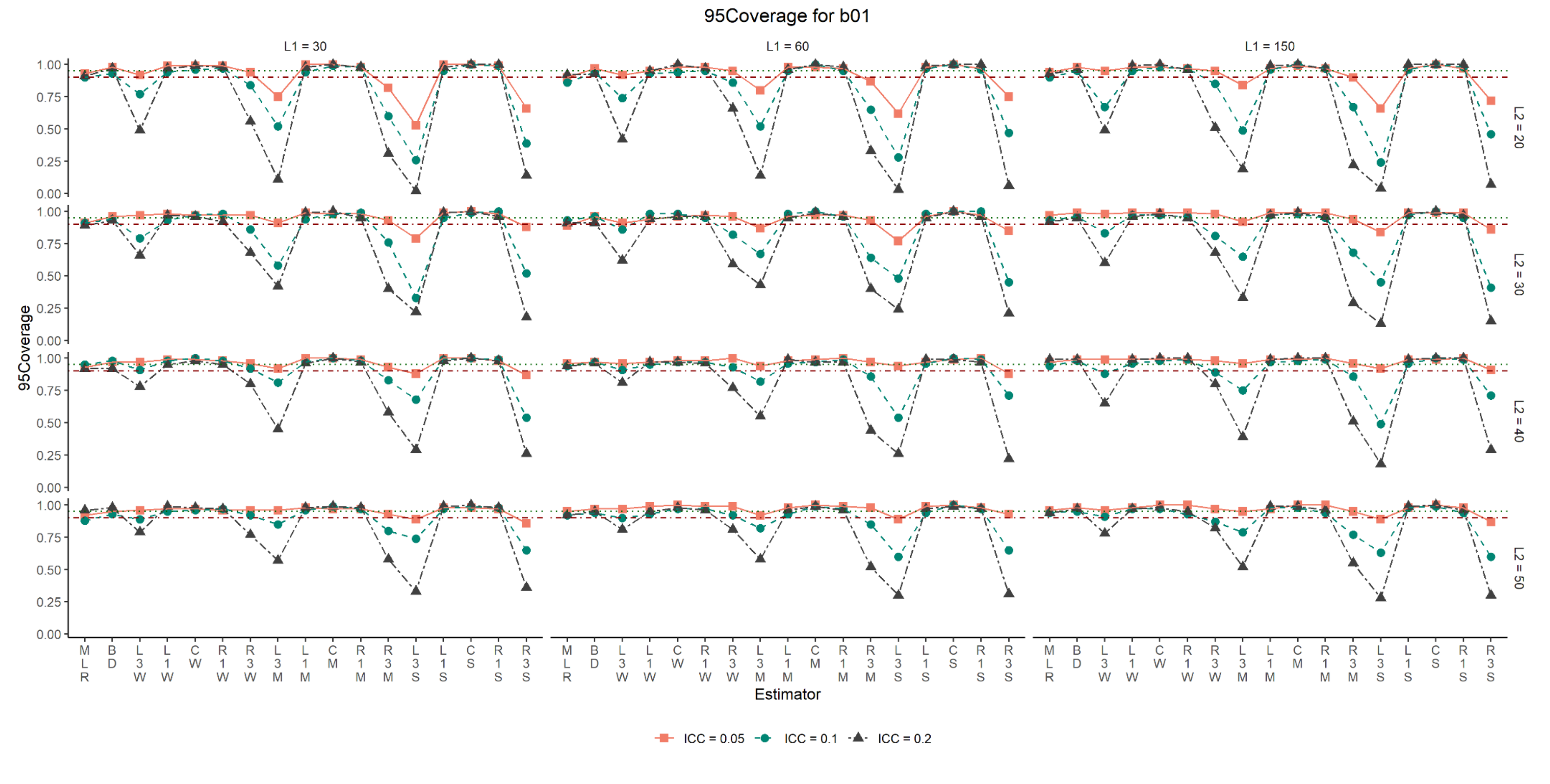



**图 7** 95% CI 覆盖率结果示例

**注：**①BD 为无信息先验的贝叶斯估计；L3 为先验分布均值向左偏离真值 3 个标准差；L1 向左偏离真值 1 个标准差；C 表示提供正确先验信息；R1 向右偏离真值 1 个标准差；R3 向右偏离真值 3 个标准差。W 表示先验信息分布的方差较大(真值的 50%)，M 表示先验信息分布的方差中等(真值的 20%)，S 表示先验信息分布的方差较小(真值的 10%)。②图中绿色虚线表示 95% CI 覆盖率为 0.95 的参照；红色点划线对应的纵坐标为 0.9，表示可接受的 95%CI 覆盖率的最小值。

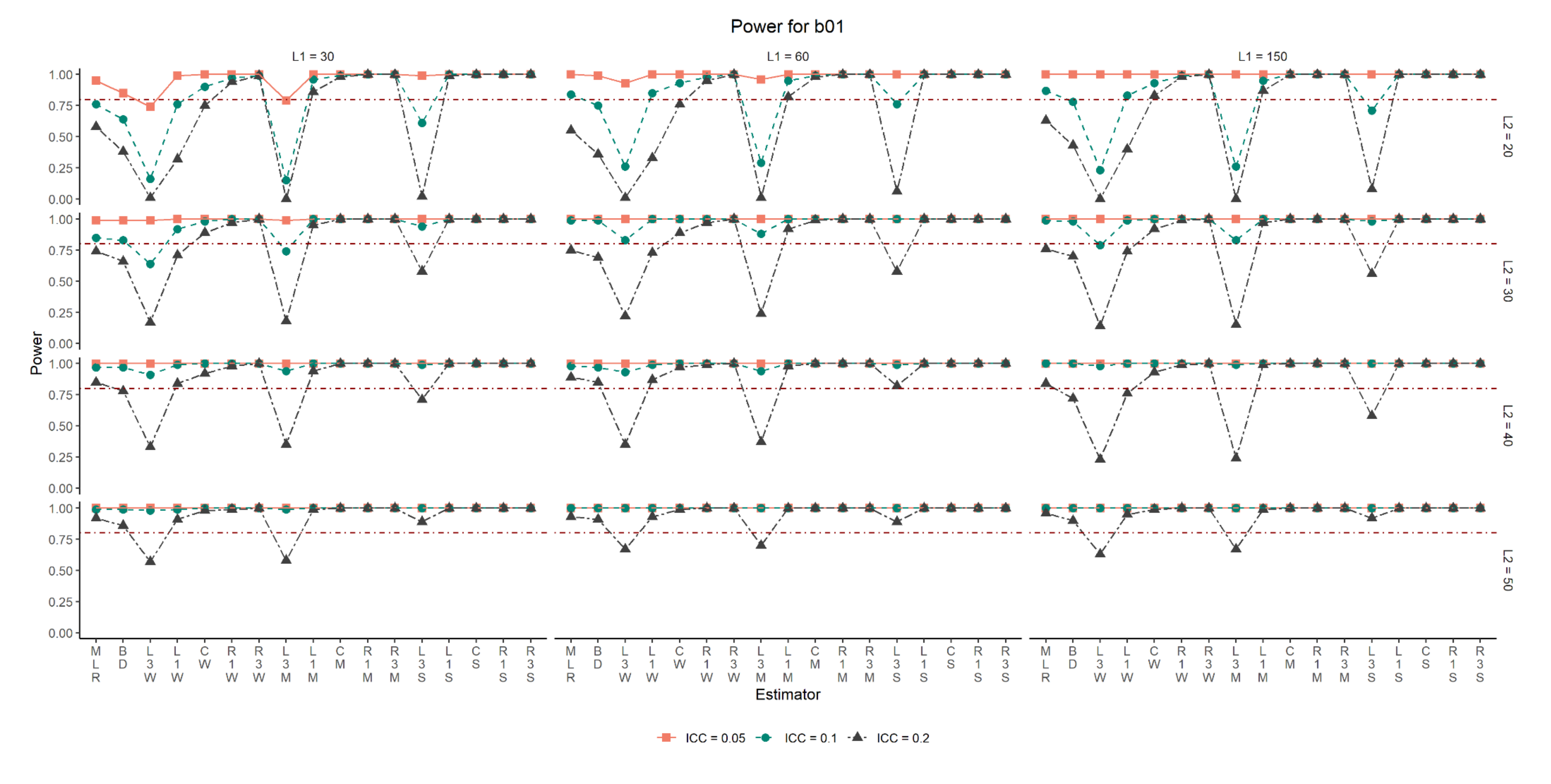



**图 8** 检验力结果示例

**注：**BD 为无信息先验的贝叶斯估计；L3 为先验分布均值向左偏离真值 3 个标准差；L1 向左偏离真值 1 个标准差；C 表示提供正确先验信息；R1 向右偏离真值 1 个标准差；R3 向右偏离真值 3 个标准差。W 表示先验信息分布的方差较大(真值的 50%)，M 表示先验信息分布的方差中等(真值的 20%)，S 表示先验信息分布的方差较小(真值的 10%)。②红色点划线对应的纵坐标为 0.8，表示可接受的检验力的最小值。

# 4 讨论

在许多心理学实验(如 cRCT)中，由于数据中可能存在的嵌套结构，研究者常常需要建立多层模型。而多层模型往往会面临样本量不足的问题，从而为研究者的数据分析造成较大的阻碍。贝叶斯估计能结合一定的先验信息，使得参数的估计更加精准，但先验分布设置的主观性可能导致错误先验信息的出现。本文基于多层模型，对不同样本量和组内相关系数条件下，错误先验信息对贝叶斯估计的影响进行研究，并给出相应的 M*plus* 代码，为贝叶斯估计中先验分布的设置提出一定的建议和指导。当 ICC 较大、水平 1 及水平 2 样本量较小时，错误先验信息对参数估计带来较大的偏差。先验分布的信息强度越强，其错误先验信息对参数估计的影响也越大。因此，在样本量较小以及 ICC 较大的情况下，需要更加警惕错误先验信息带来的影响，更谨慎对待先验分布的设置。

在模拟研究中，水平 1 及水平 2 样本量越大，错误先验信息对参数估计带来的风险也越低，这与 Holtmann 等人(2016)的结果相一致。贝叶斯方法认为未知参数是随机变量，可以利用先验信息和似然函数得到的后验分布来反映参数在不同取值上的可能性。样本量较小时，先验分布对后验分布存在较大的影响，但随着样本量的增大，先验分布的影响逐渐减弱，后验分布的形态逼近似然函数(Lynch, 2007)。因此，当样本量较小时，先验分布的错误设置往往会带来更严重的后果，需要更加谨慎。

对于 cRCT 研究中一般最受关注的效应$b_{01}$，ICC 较大时，错误先验信息对参数估计带来的负面影响较大。这与前人研究是一致的，即在多层模型中，随着 ICC 的增大，对于水平 2 效应的估计偏差随之增大，检验力随之减小(Cao et al., 2019; Finch & French, 2011)。然而需要注意的是，对于模型中的其他回归系数，ICC 对其估计偏差及检验力等方面的影响可能呈现出不一致的规律。这与 Cao 等人(2019)的研究结果同样是一致的，即对于纯粹的组内水平的效应，以及组间和组内变量的交互效应，ICC 对于其参数估计、检验力等方面的影响均没有达到显著。未来还需要通过进一步研究来更深入地探讨在多层模型中，不同水平的效应在不同 ICC 水平下的表现，以及错误先验信息对其参数估计的影响。

从不同先验分布设置的贝叶斯估计结果可以看出，先验分布的方差较大时，其错误信息带来的影响较弱。即使先验信息的设置对真实值产生一定的偏离，这种偏离带来的负面影响可以得到较好的抑制，而且随着样本量的增大，抑制的作用更明显。但当先验分布的方差较小时，即使较大的样本量也难以使得其错误先验信息的影响得到降低。因此，对参数设置方差相对较大的先验分布是更加保险的。

标准误比例方面的表现与前人研究也大致一致。随着先验分布方差的减小，标准误比例逐渐增大。这表明有信息先验的贝叶斯方法的后验分布标准差(Posterior SDs)倾向于高估参数估计实际的标准差(Empirical SDs)。但这也可能是因为随着先验分布方差的减小，参数估计的标准差也逐渐减小，从而导致标准误比例的增大(Holtmann et al., 2016)。特别是在设有错误先验信息的情况下，贝叶斯估计更倾向于高估标准误。因此，在贝叶斯估计中，研究者往往需要通过可信区间来判断参数的显著性。

## 5 局限性与展望

由于模拟研究无法穷尽所有的条件组合，本研究主要关注在 cRCT 的研究中最常见的样本量范围。然而，在实际研究中，可能出现其他样本量组合的情况。而且，本研究只考虑了不同群组内部样本量相等的情况，但实际研究中不同群组内的样本量往往很难达到平衡。未来可以进一步研究这种样本量不平衡的情况下，错误先验信息是否会带来更大的负面影响。另外，本研究在水平 1 样本量的设置上跨度较大，未来研究中可以在样本量方面进行更加精细的设置。

为了简化模拟研究的条件，本文假设因变量服从连续正态分布。然而，在实证研究中，这一假设常常会被违背。例如，由于研究者主观上有意或无意的选择倾向，或者由于天花板效应、地板效应的存在，变量数据很可能存在一定程度的偏态(Fleishman, 1978)。另外，在 cRCT 中，可能出现结果变量为分类变量，特别是二分的变量的情况(Legare et al., 2015)。因此，未来研究可以进一步探讨因变量为不同数据类型的情况下，错误先验信息可能带来的影响。

另外，考虑到 cRCT 研究中一般最受关注的是实验干预对因变量产生的效应$b_{01}$，本文主要关注的也是$b_{01}$受到不同数据条件及因素的影响。然而，对于模型中的其他回归系数，ICC 对其估计偏差及检验力等方面的影响可能呈现出不一致的规律，未来需要对多层模型中不同水平效应参数受到的来自错误先验信息的影响进行更深入的探究。

## 6 研究结论

本文聚焦于多层线性模型中不同信息强度和偏差程度的先验信息对贝叶斯估计的影响，通过进行一系列蒙特卡洛模拟，讨论了 cRCT 背景下的多层线性模型中不同样本量、ICC 和数据类型下错误先验信息对参数估计带来的影响，希望为贝叶斯估计中先验分布的设置提供一定的理论补充及实证参考。本文得出的结论如下：

(1)在小样本的多层模型中，ML 和无信息先验的贝叶斯估计可能导致有偏的组间参数估计，因此在样本量较小的情况下建议尽可能对贝叶斯方法设定一定的先验信息。

(2)严重错误的先验分布可能对贝叶斯参数估计带来毁灭性的结果，从而为实证研究带来

一定的风险。在先验分布均值偏离真值程度较小(不超过 1 个标准差)的情况下，贝叶斯估计的总体表现大体可以被接受。

(3)当水平 2 样本量及先验分布方差较小时，错误先验信息的设置会对参数估计带来较大的偏差。对于 cRCT 研究中一般最受关注的干预效应$b_{01}$，ICC 的增大也会较大程度上增大错误先验信息的负面影响。因此，在这些情况下需要更加警惕错误先验信息带来的风险。

(4)在现有先验信息不足、错误先验信息风险较大的情况下，实证研究者可以通过尽可能增大群组样本量，并适当增加先验分布的方差来降低错误先验信息可能带来的影响。



## 参考文献

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus* (Technical report, Version 4). Retrieved from http://www.statmodel.com/download/BayesAdvantages18.pdf

Aryee, S., Walumbwa, F. O., Seidu, E. Y. M., & Otaye, L. E. (2012). Impact of high-performance work systems on individual- and branch-level performance: Test of a multilevel model of intermediate linkages. *Journal of Applied Psychology, 97*, 287–300.

Campbell, M. J. (2019). Cluster randomised trials. *Medical Journal of Australia, 210*(4).

Campbell, M.J., & Walters, S.J. (2014). *How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research*. Wiley.

Cao, C., Kim, E. S., Chen, Y.-H., Ferron, J., & Stark, S. (2019). Exploring the Test of Covariate Moderation Effects in Multilevel MIMIC Models. *Educational and Psychological Measurement*, *79*(3), 512-544.

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: a comparison of four approaches. *Multivariate Behavioral Research*, *47*(6), 840-876.

Cohen, John. (1988). Statistical Power Analysis of the Behavioral Sciences ($2_{nd}$ Editor). Hillside, N. J.: Eribaum.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cornfield, J. (1978). Symposium on chd prevention trials: design issues in testing life style intervention: randomization by group: a formal analysis. *American Journal of Epidemiology, 108*(2).

Depaoli, S. (2014). The Impact of Inaccurate "Informative" Priors for Growth Parameters in Bayesian Growth Mixture Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 239-252.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian Approach to Multilevel Structural Equation Modeling With Continuous and Dichotomous Outcomes. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(3), 327-351.

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods, 22*(2).

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in crosssectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138.

Fang, J., Wen, Z. L., & Hau, K. T. (2019). Mediation Effects In 2-1-1 Multilevel Model: Evaluation Of Alternative Estimation Methods. *Structural Equation Modeling-a Multidisciplinary Journal, 26*(4), 591-606.

Finch, W. H., & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling, 18*, 229–252.

Finch, W. H., & Miller, J. E. (2019). The Use of Incorrect Informative Priors in the Estimation of MIMIC Model Parameters with Small Sample Sizes. *Structural Equation Modeling-a Multidisciplinary Journal, 26*(4), 497-508.

Fleishman, A. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521-532.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, *9*(4), 466.

Ha, A. S., Lonsdale, C., Ng, J. Y. Y., & Lubans, D. R. (2017). A school-based rope skipping program for adolescents: Results of a randomized trial. *Preventive Medicine, 101*, 188-194.

Heck, R. H. & Thomas, S. L. (2015). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus,* Routledge.

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A Comparison of ML, WLSMV, and Bayesian Methods for Multilevel Structural Equation Models in Small Samples: A Simulation Study. *Multivariate Behavioral Research, 51*(5), 661-680.

Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(2), 157–174.

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica, 64*, 157–170.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods, 6*(2), 87–93.

Kadane, J. B. (2015). Bayesian methods for prevention research. *Prevention Science, 16,* 1017–1025.

Koch, T., Schultze, M., Burrus, J., Roberts, R. D., & Eid, M. (2015). A multilevel CFA-MTMM model for nested structurally different methods. *Journal of Educational and Behavioral Statistics, 40*(5), 477–510.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*(4), 653–686.

Legare, F., Briere, N., Stacey, D., Bourassa, H., Desroches, S., Dumont, S., … Roy, L. (2015). Improving Decision making On Location of Care with the frail Elderly and their caregivers (the DOLCE study): study protocol for a cluster randomized controlled trial. *Trials, 16*.

Levy, R., & Choi, J. (2013). Bayesian structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 563–623). Charlotte, NC: Information Age.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2×2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods, 16*, 444–467.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists.* New York, NY: Springer.

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750-773.

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*. Advance online publication.

Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Incomplete draft, Version 1). Retrieved from https://pdfs.semanticscholar.org/a941/cb99e52a8eafad701d4fb0391347232779f9.pdf?_ga=2.137417013.467066868.1585219703-884182655.1568515547

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*, 599–620.

Newton, N. C., Teesson, M., Mather, M., Champion, K. E., Barrett, E. L., Stapinski, L., ... Slade, T. (2018). Universal cannabis outcomes from the Climate and Preventure (CAP) study: a cluster randomised controlled trial. *Substance Abuse Treatment Prevention and Policy, 13*.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*, 161–182.

Rutterford, Clare, Copas, Andrew, & Eldridge, Sandra. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology, 44*(3), 1051-1067.

Ribeiro, D. C., Milosavljevic, S., & Abbott, J. H. (2018). Sample size estimation for cluster randomized controlled trials. *Musculoskeletal Science and Practice, 34*, 108-111.

Ryu, E. (2015). The Role of Centering for Interaction of Level 1 Variables in Multilevel Structural Equation Models. *Structural Equation Modeling-a Multidisciplinary Journal, 22*(4), 617-630.

Schoeneberger, J. A. (2015). The impact of sample size and other factors when estimating multilevel logistic models. *The Journal of Experimental Education, 84*(3).

Shen, Y., Wang, T., Gao, M., Zhu, X., Zhang, X., He, C., ... Sun, X. (2019). Effectiveness of low-cost reminder package combined with case-based health education to improve hypertensive patients' medication adherence: a clustered randomized controlled trial. *Patient Preference and Adherence, 13*, 1083-1092.

Shi, D., & Tong, X. (2017). The impact of prior information on Bayesian latent basis growth model estimation. *SAGE Open, 7*, 1–14.

Sutton, C. J., Watkins, C. L., & Dey, P. (2013). Illustrating problems faced by stroke researchers: a review of cluster-randomized controlled trials. *International Journal of Stroke, 8*(7), 566-574.

Van Breukelen, G. J. P., & Candel, M. J. (2012). Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient!. *Journal of Clinical Epidemiology, 65*(11), 1211_2218.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*, 842– 860.

van der Putten, G.-J., Mulder, J., de Baat, C., De Visschere, L. M. J., Vanobbergen, J. N. O., & Schols, J. M. G. A. (2013). Effectiveness of supervised implementation of an oral health care guideline in care homes; a single-blinded cluster randomized controlled trial. *Clinical Oral Investigations, 17*(4), 1143-1153.

## 附录：贝叶斯多层模型 *Mplus* 语句

```
TITLE: This is an example of Bayesian Multilevel Modeling

DATA:
  FILE =data ef-nor-0.05-20-30_1.dat;

VARIABLE:
  NAMES = GroupID MemID w1j x1 x2 x3 Y;
  USEVAR = GroupID w1j x1 x2 x3 Y;
  CLUSTER = GroupID;
  WITHIN = x1 x2 x3;
  BETWEEN = w1j;

ANALYSIS:
  TYPE IS TWOLEVEL RANDOM;
  ESTIMATOR = BAYES;
  PROCESSORS = 2;
  BITERATIONS = 70000 (2000);

MODEL:
  %WITHIN%
  Y;
  b1j | Y ON x1;
  b2j | Y ON x2;
  b3j | Y ON x3;

  %BETWEEN%
  Y b1j b2j b3j;
  [Y];
  [b1j] (b10);
  [b2j] (b20);
  [b3j] (b30);
  Y ON w1j (b01);
  b1j ON w1j (b11);
  b2j ON w1j (b21);
  b3j ON w1j (b31);
  Y;
  b1j-b3j;

MODEL PRIORS:
  b01~N(1.5,0.75);
  b11~N(0.12,0.06);
  b21~N(-0.05,0.025);
  b31~N(-0.75,0.375);
```

```
  b10~N(2.3,1.15);
  b20~N(-0.25,0.125);
  b30~N(-0.75,0.375);

OUTPUT:
  TECH1 TECH8 CINT;
```